ENVIRONMENTAL
HEALTH
PERSPECTIVES

http://www.ehponline.org

A Curated Database of Rodent Uterotrophic Bioactivity

Nicole C. Kleinstreuer, Patricia C. Ceger, David G. Allen, Judy Strickland, Xiaoqing Chang, Jonathan T. Hamm, and Warren M. Casey

http://dx.doi.org/10.1289/ehp.1510183

Received: 7 May 2015
Accepted: 30 September 2015
Advance Publication: 2 October 2015

A Curated Database of Rodent Uterotrophic Bioactivity

Nicole C. Kleinstreuer[1], Patricia C. Ceger[1], David G. Allen[1], Judy Strickland[1], Xiaoqing Chang[1], Jonathan T. Hamm[1], and Warren M. Casey[2]

[1]Integrated Laboratory Systems, in support of the National Toxicology Program Interagency Center for Evaluation of Alternative Toxicological Methods (NICEATM), Research Triangle Park, North Carolina, USA; [2]National Institutes of Health/National Institute of Environmental Health Sciences/Division of the National Toxicology Program/NICEATM, Research Triangle Park, North Carolina, USA

Address correspondence to N. Kleinstreuer, P.O. Box 13501, Research Triangle Park, NC 27709 USA. Telephone: (919) 281-1110. E-mail: nicole.kleinstreuer@nih.gov

Running title: A curated rodent uterotrophic database

Acknowledgments: The authors would like to thank Kris Thayer and Mike DeVito for technical comments, Brett Jones, Eileen Phillips, and Qingda Zang for helping with the uterotrophic literature review and extracting study information, Michael Paris for assistance in formatting the uterotrophic database files, Catherine Sprankle for editorial assistance, and Steven Morefield for extracting ChemID chemical descriptor information. ILS staff provide technical support for NICEATM under NIEHS contract HHSN27320140003C, but do not represent NIEHS, NTP, or the official positions of any Federal agency.

Competing financial interests: The authors declare no competing financial interests.

Abstract

Background: Novel *in vitro* methods are being developed to identify chemicals that may interfere with estrogen receptor (ER) signaling, but results are difficult to put into biological context due to the reliance on reference chemicals established using results from other *in vitro* assays and the lack of high-quality *in vivo* reference data. The OECD-validated rodent uterotrophic bioassay is considered the "gold standard" for identifying potential ER agonists.

Objectives: We performed a comprehensive literature review to identify and evaluate data from uterotrophic studies and to analyze study variability.

Methods: We reviewed 670 articles with results from 2615 uterotrophic bioassays using 235 unique chemicals. Study descriptors, such as species/strain, route of administration, dosing regimen, lowest effect level, and test outcome, were captured in a database of uterotrophic results. Studies were assessed for adherence to six criteria based on uterotrophic regulatory test guidelines. Studies meeting all criteria (458 bioassays on 118 unique chemicals) were considered guideline-like (GL) and subsequently analyzed.

Results: The immature rat model was used for 76% of the GL studies. Active outcomes were more prevalent across rat models (74% active) compared to mouse models (36% active). Of the 70 chemicals with at least two GL studies, 18 (26%) had discordant outcomes and were classified as both active and inactive. Many discordant results were attributable to differences in study design (e.g. injection vs. oral dosing).

Conclusions: This uterotrophic database provides a valuable resource for understanding *in vivo* outcome variability and for evaluating performance of *in vitro* assays that measure estrogenic activity.

Introduction

Understanding the impact of endocrine bioactive chemicals on human health and the environment is a high priority for U.S. and international agencies. The large number of untested chemicals in commerce (>80,000) necessitates the use of high-throughput screening (HTS) programs such as the U.S. Environmental Protection Agency (EPA) ToxCast initiative and the Tox21 U.S. federal partnership to quickly identify potential endocrine disruptors and help characterize any hazards they may pose (Dix et al. 2007; Judson et al. 2010; Kavlock et al. 2012; Tice et al. 2013; U.S.EPA 2011a, 2012)Further, there is growing societal pressure to avoid animal testing and develop alternative approaches that replace, reduce, or refine the use of animals in toxicity testing (Hartung 2009; ICCVAM, 2000).

To determine the usefulness and limitations of a novel alternative method for identifying endocrine activity and to show that it is fit for its intended purpose, the method must be evaluated against a set of chemicals that have demonstrated activity and well-defined properties (potency and efficacy) against the target nuclear receptor and subsequent biological pathway. Currently, reference chemicals used to validate *in vitro* assays aimed at detecting potential endocrine disruptors (estrogen, androgen, and thyroid receptors) are selected based only on their activity in other *in vitro* assays, a circular validation paradigm that arose due to the lack of sufficient *in vivo* data (ICCVAM 2011; OECD 2012). To facilitate work that will better elucidate and characterize the relationship between *in vitro* and *in vivo* estrogen bioactivity of chemicals, the National Toxicology Program Interagency Center for Evaluation of Alternative Toxicological Methods (NICEATM) developed a curated database of high-quality *in vivo* rodent uterotrophic bioassay data extracted from published studies (http://ntp.niehs.nih.gov/pubhealth/evalatm/tox21-support/endocrine-disruptors/edhts.html).

The uterotrophic bioassay (Test Guideline [TG] 440) was validated by the Organization for Economic Co-operation and Development (OECD) as a short-term screening test to evaluate the ability of a substance to elicit estrogenic activity (Kanno et al. 2001; Kanno et al. 2003; OECD 2004; Owens and Koeter 2003). It is one of the 11 Tier 1 screening assays in the U.S. EPA's endocrine disruptor screening program (EDSP) and is considered the "gold standard" bioassay screen for identifying estrogen receptor (ER) agonists (U.S.EPA 2011b, 2012). The endpoint measured is an increase in uterine weight caused by ER-mediated water imbibition and cellular proliferation in the uterine tissue. According to the OECD ((OECD 2004)) and EPA ((U.S.EPA 2011b)) test guidelines for the uterotrophic assay, immature female rats or ovariectomized (OVX) adult female mice or rats can be used. Because immature and OVX animals do not produce endogenous estrogens, the uterus becomes sensitive to external estrogenic substances (Billon-Gales et al. 2011).

Here we describe a comprehensive database of quality-controlled *in vivo* uterotrophic studies. To create this database, we reviewed the current scientific literature for studies that measured uterine weight changes in immature rats or OVX rats or mice, identified relevant assay parameters and endpoints, compiled the data into a single database, and analyzed the data for sources of variability. Our analysis revealed that certain protocol variations, specifically use of rats versus mice and injection versus gavage dosing, were more likely to produce a positive response. This database was also used to assess the reproducibility of the uterotrophic bioassay and to provide a resource against which *in vitro* test method results for ER activity may be evaluated and predictive *in silico* models (Browne et al. 2015) may be built.

Methods

Curation process

NICEATM conducted a comprehensive literature search to identify uterotrophic studies for environmental chemicals. The ToxCast Phase I/Phase II/E1K chemical library (1812 substances, http://epa.gov/comptox/toxcast/data.html) was chosen as a starting point based on its relevance to the EDSP universe of chemicals and to facilitate future comparisons with results from the 18 HTS *in vitro* assays included in ToxCast that map to the ER pathway (Judson et al. 2015; Rotroff et al. 2014). We performed semi-automated literature searches, reviewed relevant manuscripts, and recorded detailed study information for each chemical/study/protocol combination (Table 1) along with the reported bioactivity for the dose range tested. The literature search strategy and database development procedure is illustrated in Figure 1 and detailed below.

Searches were performed in a semi-automated fashion using the U.S. National Center for Biotechnology Information's PubMatrix tool (NCBI). PubMatrix is a web-based resource that provides a simple approach to rapidly and systematically compare any list of (search) terms against any other list of (modifier) terms in PubMed. Lists of terms can include any keyword that may correspond to a Medical Subject Heading term, such as chemical names, genes, diseases, phenotypic observations, gene functions, authors, etc. Searches were performed in batches of 50 chemicals, using both chemical name and Chemical Abstracts Service Registry Number (CASRN) in the list of search terms. PubMatrix automatically identifies all chemical name synonyms in PubMed and includes these as alternative search terms. The modifier terms used to cross-reference and identify articles were “uterotrophic”, “uterotrophic assay”, and “uterine weight”. The modifier term “uterotropic” was also included as a common alternative to “uterotrophic.” The output of a PubMatrix search is a matrix table showing the frequency of co-occurrence between all pairwise comparisons between the two lists, with links out to the publications identified in the overlap space. We searched for additional studies in the U.S. Food and Drug Administration’s Endocrine Disruptor Knowledge Base (Ding et al. 2010) and the EPA’s Aggregated Computational Toxicology Resource (ACToR) database (Judson et al. 2008). Relevant publications were identified and downloaded for further manual curation, in which protocol information was entered into the NICEATM *in vivo* uterotrophic database (UTDB) so that each study could be evaluated for specifically defined quality control metrics as described below. Publications in languages other than English were included in the initial search results. These were evaluated if possible by a native language speaker but excluded from the final database of guideline-like studies.

Publications identified as measuring uterine weight changes in rats or mice were reviewed and detailed study protocol information transcribed into an Excel spreadsheet as follows. Data entry for each study protocol was performed in a standardized format and recorded in the UTDB by PubMed Identifier, CASRN, and chemical name. Two scientists independently reviewed each manuscript for relevance and extracted information on study protocol design and chemical exposure effects on uterine weight. Types of information extracted from each publication and examples are provided in Table 1. Additional information on a study protocol that did not fall into one of the predetermined study information categories was also recorded in corresponding "assay notes" and "response notes" columns. The lowest effect level (LEL), the chemical dose that caused an active outcome (a statistically significant increase in uterine weight), was reported for any compound with a positive result. The highest dose tested (HDT) was reported for chemicals with negative results. Where possible the LEL and HDT were recorded in units of mg/kg/day, although some studies reported alternate units such as mg/animal, etc. Many publications contained multiple study protocols with differing designs (e.g. comparing animal models, administration routes, or exposure durations). Pertinent details were recorded in the UTDB for every unique chemical/study protocol combination.

Study Quality Evaluation

Compliance with uterotrophic study protocol design requirements set forth in EPA OCSPP 890.1600 (U.S.EPA 2011b) and OECD TG 440 (OECD 2004) was evaluated based on the information extracted from each publication. Two scientists independently scored each protocol for adherence to six predefined minimum criteria (MC) for a guideline-like study. A study protocol was considered to be "guideline-like" (GL) if all six of the MC shown in Figure 2 and explained in the following paragraph were met.

Acceptable animal models included immature rats, OVX adult rats, and OVX adult mice. Based on OECD recommendations, studies using immature mice were not considered to be GL due to their potential insensitivity to weak estrogens (OECD 2004). For studies using the OVX animal model, we required the ovariectomy to have been performed between six and eight weeks of age, allowing at least 14 days post-surgery before dosing for rats and seven days post-surgery for mice to ensure adequate time for uterine tissues to regress. For immature rat studies, the dosing should have begun after weaning between postnatal day (PND) 18 and PND 21, and been completed by PND 25 (before the onset of puberty). Each positive or negative control group was required to have a minimum of three animals, and each test group was required to have a minimum of five animals. This requirement differs from the OECD and EPA guidelines, both of which require six animals in both control and test groups (OECD 2004; U.S.EPA 2011b). However, we found that a large number of studies using marginally smaller group sizes fulfilled every other MC to be considered GL, so we relaxed these criteria to be slightly more inclusive while still ensuring sufficient statistical power. Acceptable routes of administration included oral gavage (p.o.), subcutaneous (s.c.) injection, and intraperitoneal (i.p.) injection, although both the OECD and EPA guidelines state that injection routes are preferred to increase bioavailability of the test substance. We required a minimum of two dose groups treated over a minimum dosing interval of three consecutive days to show dose-dependent effects and establish a LEL. Finally, to ensure appropriate timing for effect evaluation, we required the necropsy to have been carried out 18–36 hours after the last dose. Compared to the OECD and EPA guidelines, which specify that necropsy should occur 24 hours after the last dose (OECD 2004; U.S.EPA 2011b), this requirement was expanded to maximize the number of adherent studies. We recorded data indicating whether or not levels of phytoestrogen in the diet were reported, but this criterion was not incorporated into the final GL criteria due to the small number of studies reporting this information (<5% of the 670 papers reviewed).

A score of 0 (no) or 1 (yes) was recorded for each of the minimum criteria (MC 1-6) based on whether the study protocol fulfilled that particular requirement. These scores were recorded as individual columns in the UTDB and added to yield a total score for each study protocol. The two independent evaluations for each study protocol were compared. If the two evaluations concurred, information from that study protocol was entered into the final version of the UTDB. If the two evaluations differed, the paper was re-reviewed to identify the source of the discrepancy and reach a consensus. Only those protocols meeting all six criteria were considered GL. The subset of GL uterotrophic study protocols constitutes the GL uterotrophic database (GL-UTDB).

It should be noted that compliance with the MC identified above is not necessarily equivalent to a thorough assessment of overall study quality. For example, our evaluation did not consider the internal validity of each study, risk of bias, or whether the route of administration was relevant to the expected route of human exposure.

Results

The search for uterotrophic data on the 1812 ToxCast compounds (http://epa.gov/comptox/toxcast/data.html) yielded over 1000 papers, of which 670 were deemed potentially relevant based on the inclusion of uterine weight as a measured endpoint. From these 670 manuscripts, 2615 individual chemical/study/protocol combinations were extracted, yielding results on 235 chemicals with unique CASRNs (http://ntp.niehs.nih.gov/pubhealth/evalatm/tox21-support/endocrine-disruptors/edhts.html). It was common for one paper to contain multiple study design protocols, of which only some protocols met all six MC and were included in the GL-UTDB (http://ntp.niehs.nih.gov/pubhealth/evalatm/tox21-support/endocrine-disruptors/edhts.html). The GL-UTDB contains information from 458 GL studies extracted from 93 publications, providing high-quality *in vivo* estrogenic bioactivity data on 118 chemicals with unique CASRNs (103 of which are in the ToxCast/Tox21 inventory). We included all chemicals in the studies returned by our search, some of which were not in the ToxCast library but were included in publications that also examined ToxCast chemicals. We performed an additional round of manual quality assurance on all study information in the GL-UTDB to confirm data entry accuracy. To facilitate computational analyses, we added standardized chemical descriptor information (ChemID number, ChemID name, and molecular formula, available via http://chem.sis.nlm.nih.gov/chemidplus/) and a "protocol" variable that computationally binds multiple fields together to provide a unique identifier for each study.

Impact of study design on uterotrophic outcome

Six basic study designs met GL criteria, depending on species (rat or mouse), route of administration (oral or injection), and use of OVX (rat or mouse) or immature (rat only) animals. The majority of studies that met GL criteria were performed with either s.c. or i.p. routes of injection (69% [317/458]). Both injection routes are acceptable according to OECD and EPA guidelines (OECD 2004; U.S.EPA 2011b) and thus, for this analysis, "injection" refers to studies using either s.c. or i.p. routes of administration. However, it should be noted that 99% (313/317) of the injection studies in the database used the s.c. route.

A breakdown of results by study design is provided in Table 2. Data from two chemicals commonly used as positive controls (ethinyl estradiol and estradiol) were excluded from this analysis due to the large number of results and inherent bias associated with their inclusion (*i.e.*, negative results would indicate a failed "positive" control and would therefore not typically be reported), leaving 374 GL uterotrophic entries. The immature rat model was used for 76% (285/374) of the studies in the database, with 72% (204/285) of these studies using injection as the route of administration. Active outcomes were more prevalent in rat models (74% [242/327] of all rat outcomes were active). This is in contrast to uterotrophic results reported in mouse models, in which 36% (17/47) of all outcomes were active, with the OVX_mouse_oral design producing active outcomes in only 27% (6/22) of the studies. It should be noted that the selection of chemicals tested in these studies is neither random nor uniformly distributed with respect to uterotrophic bioactivity, and the performance of a particular study protocol design, especially those with a small number of examples (e.g., OVX_rat_injection or OVX_mouse_oral), could be heavily influenced by a single publication from one laboratory testing multiple chemicals in that particular study design.

Reproducibility of uterotrophic outcomes

The GL-UTDB provides an opportunity to assess both the qualitative and quantitative reproducibility of the uterotrophic assay across many chemicals tested at many different laboratories. Of the 70 chemicals in the database with at least two reported GL uterotrophic studies (Figure 3**)**, 18 (26%) had at least one study with a discordant outcome, resulting in a chemical being classified as both "active" and "inactive" for uterotrophic bioactivity. Table 3 lists chemicals with discordant results along with the minimum reported LEL and maximum reported highest dose tested (HDT) for each chemical. Discordant outcomes could result both from differences in overall study protocol design and/or the range of doses tested in each study. For example, the HDT from an inactive result may have been below the dose that would produce a tissue concentration required for bioactivity, as appears to be the case for benzephenone, permethrin, and daidzein. In other cases, the HDT for an inactive result may have been very close or equal to the minimum LEL (minLEL) for an active result, and discrepancies could be attributed to small increases that either just crossed the threshold or failed to reach statistical significance. This was observed for diethylstilbestrol, a known estrogenic compound, where a dose of 0.05 µg/kg/day produced a ~1.3 fold increase in uterine weight (p<0.01) in one study (Odum et al. 2002) and produced a non-statistically significant increase of ~20% at the same dose in a different study (Tinwell and Ashby 2004), both with the same basic study design. However, in that same paper (Tinwell and Ashby 2004) with the inactive result, there were additional experimental protocols that showed significant uterotrophic activity at slightly higher doses of 0.25 µg/kg/day. The GL-UTDB contains one additional compound, 4-nonylphenol (branched form, CASRN: 25154-52-3), that had 22 active results (minLEL of 5 mg/kg/day) and 2 inactive results (maximum HDT [maxHDT] of 80 mg/kg/day), but this was found to be a mixture of branched chains rather than a unique structure. Because we could not ascertain that the same form was being tested in each study, it was excluded from this analysis.

Of the 18 chemicals listed in Table 3, 10 (56%, shaded rows in the table) had discordant uterotrophic outcomes that may be attributable to differences in study protocol design. Results from the testing of butylparaben provide an example of how study design can impact uterotrophic outcomes, as shown in the radar plot in Figure 4**.** In this case, all eight active results were reported in the three study protocol designs using s.c. injection as the route of administration (immature rat, OVX rat, OVX mouse), whereas inactive results were reported for both study protocol designs that used oral dosing (immature rat, OVX mouse). In all three injection protocols, the minLEL reported was well below the maximum highest dose tested in the oral dosing protocols. A similar radar plot for each chemical in Table 3, illustrating the relationship between study protocol design and outcome, is provided in Supplemental Material, Figure S1.

The eight chemicals in non-shaded rows in Table 3 had discordant outcomes reported for studies using the same basic study design. Uterotrophic outcomes were compared to determine if the HDT for inactive outcomes was below the LEL reported for active outcomes, in which case the results would actually support one another. For chemicals that had discordant outcomes reported in the same study design, it was common for the HDT to be above LEL doses reported in other studies, although the difference in these values was normally less than one order of magnitude. Most studies in the UTDB and the GL-UTDB typically used no more than four log-spaced doses, resulting in poor resolution of lowest effect levels (generally defined as >20% increase in wet uterine weight, $p<0.05$), which could explain LELs and HDT reported at similar doses. However, reports of inactive results obtained at doses well above all reported LELs are difficult to reconcile. Figure 5 shows discordant results for chemicals tested using the same basic study design: immature rat, s.c. injection, which was the most common design and correspondingly had the highest number of discrepancies. Bisphenol A (BPA, CASRN 80-05-7) provides a good example of the high degree of variability that can be seen in the uterotrophic bioassay, with BPA classified as “active” by s.c. injection at 2 mg/kg/day in one study using the immature rat model (CERI-METI 2006), and “inactive” by s.c. injection at 1000 mg/kg/day in another study using the same model (An et al. 2002).

Chemicals with independently reproducible uterotrophic outcomes

There were 36 chemicals (24 active, 12 inactive) that showed reproducible results in two or more independent GL uterotrophic studies (Table 4**)**. The minLEL and maxHDT are given in mg/kg/day; however, this information cannot necessarily be translated into expected potency values as it is inherently limited for some compounds by the dose ranges selected in the studies. Further, there are studies with potentially lower LELs than those reported in Table 4 that were given in terms of mg/animal/day or total dose. For consistency we used the minLEL from studies that reported units of mg/kg/day, unless the only studies reporting outcomes for that chemical reported doses in units other than mg/kg/day.

The active compounds included steroid pharmaceuticals commonly used as positive controls and multiple BPA analogues, while the inactive compounds included several phthalates. Also included in the actives list are two well-known selective estrogen receptor modulators with both agonist and antagonist activities, tamoxifen and clomiphene citrate (Mirkin and Pickar 2015). There were two additional active compounds (gibberellic acid and tiratricol) with LELs in more than one protocol, but they were part of the same study by the same lab and so were not considered independently reproduced. Similarly, there were thirteen inactive compounds that were negative in multiple protocols run as part of one study, and thus are not shown in Table 4. Ten of these thirteen were from a study that was part of the OECD validation that examined both s.c. and p.o. routes of administration in immature rats (Ohta et al. 2012).

Discussion

U.S. and international regulations require the testing of chemicals for the detection of potential endocrine disruptors, but there are thousands of chemicals in commerce for which no data are currently available. *In vitro* HTS screening assays have been developed to fill some of these data gaps in a timely and cost-effective manner, but in order to use these data for hazard identification purposes, the usefulness and limitations of these *in vitro* assays must be carefully evaluated. To better understand and characterize the relationship between *in vitro* and *in vivo* activity of potential endocrine disruptors, we developed a curated database of high-quality *in vivo* data found in the literature relevant to estrogen receptor agonism. We focused specifically on the estrogen receptor pathway because of the large number of chemicals that have been tested in the uterotrophic assay, an *in vivo* screening test that has undergone international validation by OECD (Kanno et al. 2001, 2003; Owens and Koeter 2003) and is included in the EPA's EDSP Tier 1 battery (U.S.EPA 2012).

Our database and accompanying analyses and chemical lists represent the first of at least three such efforts describing the *in vivo* endocrine activity of chemicals encompassing the estrogen, androgen, and thyroid pathways, respectively. This curated information serves as a valuable anchoring point for assessing the impact of study design on test results, the reproducibility of chemical activity, and the performance of *in vitro*/computational approaches. We have provided herein a transparent outline of the strategies used to identify rodent uterotrophic studies. Data were extracted from the literature, reviewed by two independent reviewers, and assigned a score based on minimum criteria derived to mimic the study parameters defined in EPA and OECD test guidelines accepted by U.S. and international regulatory authorities. In total, more than 40 parameters were extracted from each study to allow downstream analyses of their relative impact on study results. The large number of chemicals included in the GL-UTDB far exceeds the seven chemicals examined in the OECD validation of the uterotrophic assay (OECD 2007), and may provide a more robust assessment of the experimental variability associated with this *in vivo* test method.

Our results reveal substantive variability in the *in vivo* outcomes for chemicals tested more than once, which will be valuable information in characterizing the relevance and reliability of proposed alternatives. We analyzed sources of variability in outcomes and study designs and found that these discordances were largely attributed to differences in study design, which were most often based on differences in dosing route or maximum dose tested. The substantially higher number of positive outcomes in injection studies as compared to oral studies highlights the need to understand the impact of exposure route and metabolism on actual tissue dose, and to employ reverse dosimetry to more accurately extrapolate from *in vitro* to *in vivo* bioactivity (Chang et al. 2014; Wetmore et al. 2012; Wetmore 2015). When establishing performance metrics for any alternative test method, it is important to consider both the inherent variability of the *in vivo* method and variability associated with using different protocols. For example, inherent variability has been attributed to potential false negatives in the uterotrophic assay due to the limited number of animals used in each group, the relatively short duration of a study, and the variability in control uterus weights (Ashby and Odum 2004; Christian et al. 1998). An alternative method, such as the ToxCast assays, may realistically be expected to predict the true response but not necessarily the associated *in vivo* experimental variability (Browne et al. 2015).

We have focused on high-quality studies that met all our minimum criteria to be considered GL. However, we have included all the necessary information for others to re-analyze the data in a more inclusive or more stringent fashion as fits their needs, whether those needs are research- or regulatory-related. There are undoubtedly a number of reliable studies in the UTDB that did not meet all six minimum criteria whose data could be included in future analyses; these include positive results from assays performed in immature mice (Ding et al. 2010; Hossaini et al. 2000; Tinwell et al. 2000) or single dose studies that were part of the OECD validation (Kim et al. 2005).

Conclusion

We anticipate that the uterotrophic results compiled for this manuscript will serve as a valuable resource for understanding sources of *in vivo* study variability and reproducibility, providing biological context for data generated from *in vitro* estrogen receptor agonist assays, and anchoring predictive *in silico* models for estrogenic bioactivity via identification of estrogen agonist reference chemicals.

References

An BS, Kang SK, Shin JH, Jeung EB. 2002. Stimulation of calbindin-d(9k) mrna expression in the rat uterus by octyl-phenol, nonylphenol and bisphenol. Molecular and cellular endocrinology 191:177-186.

Ashby J, Odum J. 2004. Gene expression changes in the immature rat uterus: Effects of uterotrophic and sub-uterotrophic doses of bisphenol a. Toxicological Sciences 82:458-467.

Billon-Gales A, Krust A, Fontaine C, Abot A, Flouriot G, Toutain C, et al. 2011. Activation function 2 (af2) of estrogen receptor-alpha is required for the atheroprotective action of estradiol but not to accelerate endothelial healing. Proceedings of the National Academy of Sciences of the United States of America 108:13311-13316.

Browne P, Judson R, Casey W, Kleinstreuer N, Thomas R. 2015. Screening chemicals for estrogen receptor bioactivity using a computational model. Environmental science & technology 49:8804-8814.

CERI-METI. 2006. Draft report of pre-validation and inter-laboratory validation for stably transfected transcriptional activation (ta) assay to detect estrogenic activity (http://www.Oecd.Org/chemicalsafety/testing/37504278.Pdf). Tokyo, Japan.

Chang X, Kleinstreuer N, Ceger P, Hsieh J-H, Allen D, Casey W. 2014. Application of reverse dosimetry to compare in vitro and in vivo estrogen receptor activity. Applied In Vitro Toxicology 1:33-44.

Christian MS, Hoberman AM, Bachmann S, Hellwig J. 1998. Variability in the uterotrophic response assay (an in vivo estrogenic response assay) in untreated control and positive control (des-dp, 2.5 microg/kg, bid) wistar and sprague-dawley rats. . Drug and Chemical Toxicology 21 Supple 1:50.

Ding D, Xu L, Fang H, Hong H, Perkins R, Harris S, et al. 2010. The edkb: An established knowledge base for endocrine disrupting chemicals. BMC bioinformatics 11 Suppl 6:S5.

Dix DJ, Houck KA, Martin MT, Richard AM, Setzer RW, Kavlock RJ. 2007. The toxcast program for prioritizing toxicity testing of environmental chemicals. Toxicological sciences : an official journal of the Society of Toxicology 95:5-12.

Hartung T. 2009. Toxicology for the twenty-first century. Nature 460:208-212.

Hossaini A, Larsen JJ, Larsen JC. 2000. Lack of oestrogenic effects of food preservatives (parabens) in uterotrophic assays. Food and chemical toxicology : an international journal published for the British Industrial Biological Research Association 38:319-323.

ICCVAM. Iccvam authorization act of 2000. H.R. 4281. Public Law 106-545.

ICCVAM. 2000 Iccvam authorization act of 2000. H.R. 4281. Public Law 106-545.

ICCVAM. 2011. Test method evaluation report: The lumi-cell (bg1 luc er ta) test method: An in vitro assay for identifying human estrogen receptor agonist and antagonist activity of chemicals. Online: http://ntp.niehs.nih.gov/iccvam/docs/endo_docs/erta-tmer/bg1lucer-ta-tmer-combined.pdf.

Judson R, Richard A, Dix D, Houck K, Elloumi F, Martin M, et al. 2008. Actor--aggregated computational toxicology resource. Toxicology and applied pharmacology 233:7-13.

Judson RS, Houck KA, Kavlock RJ, Knudsen TB, Martin MT, Mortensen HM, et al. 2010. In vitro screening of environmental chemicals for targeted testing prioritization: The toxcast project. Environmental health perspectives 118:485-492.

Judson RS, Magpantay FM, Chickarmane V, Haskell C, Tania N, Taylor J, et al. 2015. Integrated model of chemical perturbations of a biological pathway using 18 in vitro high throughput screening assays for the estrogen receptor. Toxicological Sciences.

Kanno J, Onyon L, Haseman J, Fenner-Crisp P, Ashby J, Owens W, et al. 2001. The oecd program to validate the rat uterotrophic bioassay to screen compounds for in vivo estrogenic responses: Phase 1. Environmental health perspectives 109:785-794.

Kanno J, Onyon L, Peddada S, Ashby J, Jacob E, Owens W. 2003. The oecd program to validate the rat uterotrophic bioassay. Phase 2: Dose-response studies. Environmental health perspectives 111:1530-1549.

Kavlock R, Chandler K, Houck K, Hunter S, Judson R, Kleinstreuer N, et al. 2012. Update on epa's toxcast program: Providing high throughput decision support tools for chemical risk management. Chemical research in toxicology 25:1287-1302.

Kim HS, Kang TS, Kang IH, Kim TS, Moon HJ, Kim IY, et al. 2005. Validation study of oecd rodent uterotrophic assay for the assessment of estrogenic activity in sprague-dawley immature female rats. Journal of toxicology and environmental health Part A 68:2249-2262.

Mirkin S, Pickar JH. 2015. Selective estrogen receptor modulators (serms): A review of clinical data. Maturitas 80:52-57.

NCBI. Pubmatrix: Http://pubmatrix.Grc.Nia.Nih.Gov/ [accessed August 2013 - December 2014.

Odum J, Lefevre PA, Tinwell H, Van Miller JP, Joiner RL, Chapin RE, et al. 2002. Comparison of the developmental and reproductive toxicity of diethylstilbestrol administered to rats in utero, lactationally, preweaning, or postweaning. Toxicological sciences : an official journal of the Society of Toxicology 68:147-163.

OECD. 2004. Test no. 440. Uterotrophic bioassay in rodents: A short-term screening test for oestrogenic properties. Oecd guidelines for the testing of chemicals, section 4: Health effects. . Online: http://www.oecd-ilibrary.org/environment/test-no-440-uterotrophic-bioassay-in-rodents_9789264067417-en.

OECD. 2007. Series on testing and assessment. Number 67. Report of the validation of the uterotrophic bioassay: Additional data supporting the test guideline on the uterotrophic bioassay in rodents. . Available: http://epa.gov/endo/pubs/uterotrophic_OECD_validation_report.pdf.

OECD. 2012. Test no. 455: Performance-based test guideline for stably transfected transactivation in vitro assays to detect estrogen receptor agonists. Online: http://www.oecd-ilibrary.org/environment/test-no-455-performance-based-test-guideline-for-stably-transfected-transactivation-in-vitro-assays-to-detect-estrogen-receptor-agonists_9789264185388-en.

Ohta R, Takagi A, Ohmukai H, Marumo H, Ono A, Matsushima Y, et al. 2012. Ovariectomized mouse uterotrophic assay of 36 chemicals. The Journal of toxicological sciences 37:879-889.

Owens W, Koeter HB. 2003. The oecd program to validate the rat uterotrophic bioassay: An overview. Environmental health perspectives 111:1527-1529.

Rotroff DM, Martin MT, Dix DJ, Filer DL, Houck KA, Knudsen TB, et al. 2014. Predictive endocrine testing in the 21st century using in vitro assays of estrogen receptor signaling responses. Environmental science & technology 48:8706-8716.

Tice RR, Austin CP, Kavlock RJ, Bucher JR. 2013. Improving the human hazard characterization of chemicals: A tox21 update. Environmental health perspectives 121:756-765.

Tinwell H, Joiner R, Pate I, Soames A, Foster J, Ashby J. 2000. Uterotrophic activity of bisphenol a in the immature mouse. Regulatory toxicology and pharmacology : RTP 32:118-126.

Tinwell H, Ashby J. 2004. Sensitivity of the immature rat uterotrophic assay to mixtures of estrogens. Environmental health perspectives 112:575-582.

U.S.EPA. 2011a. The incorporation of in silico models and in vitro high throughput assays in the endocrine disruptor screening program (edsp) for prioritization and screening. Summary overview. A part of the edsp comprehensive management plan Online: http://epa.gov/endo/pubs/edsp21_work_plan_summary%20_overview_final.pdf

U.S.EPA. 2011b. Standard evaluation procedure uterotrophic assay ocspp 890.1600. Online: http://www.epa.gov/endo/pubs/toresources/seps/Final_890.1600_Uterotrophic_Assay_SEP%209.22.11.pdf.

U.S.EPA. 2012. Endocrine disruptor screening program universe of chemicals and general validation principles. Online:

http://www.epa.gov/endo/pubs/edsp_chemical_universe_and_general_validations_white_paper_11_11.pdf.

Wetmore BA, Wambaugh JF, Ferguson SS, Sochaski MA, Rotroff DM, Freeman K, et al. 2012. Integration of dosimetry, exposure, and high-throughput screening data in chemical toxicity assessment. Toxicological sciences : an official journal of the Society of Toxicology 125:157-174.

Wetmore BA. 2015. Quantitative in vitro-to-in vivo extrapolation in a high-throughput environment. Toxicology 332:94-101.

Table 1. Study details (and examples) extracted from papers measuring uterine weight change.

Study Information Category	**Examples[a]**
Species	Rat, Mouse
Strain	Sprague Dawley, Wistar, CD1, etc.
Study Type	Immature, Ovariectomized, Intact, etc.
Assay Type	Organ weight
Assay Target	Uterine weight
Route of Administration	i.p. injection, s.c. injection, P.O., etc.
Age at First Dose	PND 0, PND 18, Adult, etc.
OVX Status	OVX or NA
Age at OVX	PND 20, 5 weeks, NA, etc.
Dosing Length	Single dose, 3 days, 3 weeks, etc.
Dosing Frequency	Daily, Twice daily, etc.
Number of Doses	1, 2, 3, 4, etc.
Highest Dose Tested	500 mg/kg/day, etc.
Number of Animals	3, 4, 5, 6, etc.
Positive Control	Estradiol, Ethinyl estradiol
Post-Treatment Necropsy Time	24 hr, 1 day, etc.
Lowest Effect Level (LEL)	0.1, 10, 100, etc.
LEL Units	mg/kg/day, mg/animal, etc.
Response Observed	Increase, Decrease, NA
Response Value	1.5, 2; 150, 200; 0.01, 0.2; etc.
Response Units	Fold Change relative to control; Percent Increase; Log Relative Potency; etc.

Abbreviations: OVX: ovariectomized, i.p.: intraperitoneal, s.c.: subcutaneous, p.o.: oral gavage, PND: postnatal day, NA: not available.

[a] Examples for response units correspond to the types of response values collected.

Table 2. Distribution of uterotrophic outcomes by study design (GL studies only).

Outcome	Imm_Rat Inj	Imm_Rat Oral	OVX_Rat Inj	OVX_Rat Oral	OVX_Mouse Inj	OVX_Mouse Oral
# Active[a]	147	61	29	5	11	6
# Inactive	57	20	3	5	14	16
% Active	0.72	0.75	0.91	0.50	0.44	0.27
% Inactive	0.28	0.25	0.09	0.50	0.56	0.73
% Total	0.545	0.217	0.086	0.027	0.067	0.059

Abbreviations: GL: guideline-like, Imm: immature, OVX: ovariectomized, Inj: injection (either subcutaneous or intraperitoneal), Oral: oral gavage. # Active: the number of experiments reporting substances as active, # Inactive: the number of experiments reporting substances as inactive.

[a] Data for positive controls are not included in this table.

Table 3. Chemicals with discordant uterotrophic results in GL studies.

CASRN	Name[a]	# GL Active	minLEL (mg/kg/day)	# GL Inactive	maxHDT (mg/kg/day)
80-05-7	Bisphenol A	37	2	6	1000
446-72-0	Genistein	27	1	1	5
72-43-5	Methoxychlor	18	20	1	200
789-02-6	*o,p'*-DDT	15	1	1	100
94-26-8	Butylparaben	8	50	2	1000
56-53-1	Diethylstilbestrol	8	0.00005	1	0.00005
104-40-5	4-n-Nonylphenol (linear, *para*)	5	75	4	200
140-66-9	4-(1,1,3,3-Tetramethylbutyl)phenol	3	56	1	250
120-47-8	Ethylparaben	1	180	3	1000
119-61-9	Benzophenone	1	500	2	200
99-76-3	Methylparaben	1	55	2	800
56-55-3	Benz(a)anthracene	1	1	2	300
1806-26-4	4-Octylphenol	1	100	2	200
94-13-3	Propylparaben	1	65	2	1000
52645-53-1	Permethrin	1	800	1	150
50-55-5	Reserpine	1	3	1	3
520-36-5	Apigenin	1	5	1	200
486-66-8	Daidzein	1	600	1	200

Abbreviations: GL: guideline-like; CASRN: Chemical Abstracts Service Registry Number; minLEL: minimum lowest effect level; maxHDT: maximum highest dose tested.

[a] Shaded chemicals had discordant uterotrophic outcomes in guideline like (GL) study designs that differed significantly from one another, and non-shaded chemicals had discordant results reported in assays with the same basic study design.

Table 4. Chemicals with independently reproduced concordant guideline-like (GL) uterotrophic results.

CASRN	Name	GL Active	GL Inactive	Bioactivity	minLEL (mg/kg/day)[a]	maxHDT (mg/kg/day)
50-28-2	Estradiol	25	0	Active	0.00001	NA
57-63-6	Ethinyl Estradiol	59	0	Active	0.0001	NA
72-33-3	Mestranol	3	0	Active	0.00008*	NA
50-27-1	Estriol	4	0	Active	0.002*	NA
10540-29-1	Tamoxifen	12	0	Active	0.01	NA
57-91-0	Alfatradiol	2	0	Active	0.4	NA
68-22-4	Norethindrone	2	0	Active	2	NA
53-16-7	Estrone	9	0	Active	2	NA
474-86-2	Equilin	2	0	Active	2	NA
17924-92-4	Zearalenone	4	0	Active	2	NA
50-41-9	Clomiphene citrate	2	0	Active	2	NA
1478-61-1	Bisphenol AF	4	0	Active	4	NA
58-18-4	Methyltestosterone	3	0	Active	10	NA
80-09-1	Bisphenol S	2	0	Active	20	NA
77-40-7	Bisphenol B	2	0	Active	20	NA
599-64-4	p-Cumylphenol	2	0	Active	20	NA
521-18-6	Dihydrotestosterone	3	0	Active	20	NA
104-43-8	4-Dodecylphenol	3	0	Active	40	NA
98-54-4	p-tert-Butylphenol	2	0	Active	100	NA
131-56-6	2,4-Dihydroxybenzophenone	3	0	Active	100	NA
80-46-6	4-(1,1-Dimethylpropyl)phenol	4	0	Active	200	NA
5153-25-3	Benzoic acid, 4-hydroxy-, 2-ethylhexyl ester	2	0	Active	200	NA
131-55-5	Benzophenone-2	6	0	Active	200	NA
556-67-2	Octamethylcyclotetrasiloxane	3	0	Active	250	NA
51630-58-1	Fenvalerate	0	2	Inactive	NA	80
1461-22-9	Tributylchlorostannane	0	2	Inactive	NA	200
99-96-7	4-Hydroxybenzoic acid	0	2	Inactive	NA	1000
87-86-5	Pentachlorophenol	0	2	Inactive	NA	1000
84-75-3	Dihexyl phthalate	0	2	Inactive	NA	1000
84-74-2	Dibutyl phthalate	0	2	Inactive	NA	1000
84-61-7	Dicyclohexyl phthalate	0	2	Inactive	NA	1000
61-82-5	Amitrole	0	2	Inactive	NA	1000
520-18-3	Kaempferol	0	3	Inactive	NA	1000
117-81-7	Bis(2-ethylhexyl)phthalate	0	2	Inactive	NA	1000
103-23-1	Bis(2-ethylhexyl)hexanedioate	0	2	Inactive	NA	1000
84-66-2	Diethyl phthalate	0	2	Inactive	NA	2000

Abbreviations: GL: guideline-like; CASRN: Chemical Abstracts Service Registry Number; minLEL: minimum lowest effect level; maxHDT: maximum highest dose tested; NA: not applicable.

[a] The minLEL (for active chemicals) and maxHDT (for inactive chemicals) are shown in units of mg/kg/day, except in the cases of mestranol and estriol, where the only reported minLELs were in mg/rat/day (annotated by *).

Figure Legends

Figure 1. Flow diagram illustrating the curation of the uterotrophic database (UTDB) and identification of high-quality guideline like (GL) studies. Abbreviations: EDKB: Endocrine Disruptor Knowledge Base, ACToR: Aggregated Computational Toxicology Resource, LEL: lowest effect level.

Figure 2. Minimum criteria for guideline-like (GL) uterotrophic studies. Abbreviations: OVX: ovariectomized, PND: postnatal day.

Figure 3. Results from uterotrophic studies for chemicals that had at least two independent GL studies. Dark gray bars represent the number of "active" reports; light gray bars represent the number of inactive reports. Data from chemicals commonly used as positive controls (i.e., ethinyl estradiol and estradiol) were excluded from this plot.

Figure 4. Example of butylparaben, where differences in study protocol design that may be associated with discordant uterotrophic outcomes. Numbers of active (black) and inactive (gray) outcomes are shown (dotted lines represent number of outcomes, maximum of 5 here) for butylparaben as a function of study design. The minimum lowest effect level (minLEL) is reported for the 8 active outcomes (5 Imm_rat_inj, 2 OVX_mouse_inj, 1 OVX_rat_inj) and the maximum highest dose tested (maxHDT) is reported for the 2 inactive outcomes (1 Imm_rat_oral, 1 OVX_mouse_oral). Abbreviations: Imm: immature, OVX: ovariectomized, inj: injection (either subcutaneous or intraperitoneal), oral: oral gavage.

Figure 5. LELs and HDTs for six chemicals with discordant results in the Immature_Rat_Injection study design. Markers reflect lowest effect levels (LEL) for chemicals classified as "active" in the uterotrophic bioassay (light gray markers), and highest dose tested (HDT) for those with "inactive" uterotrophic outcomes (dark gray markers).

Figure 1

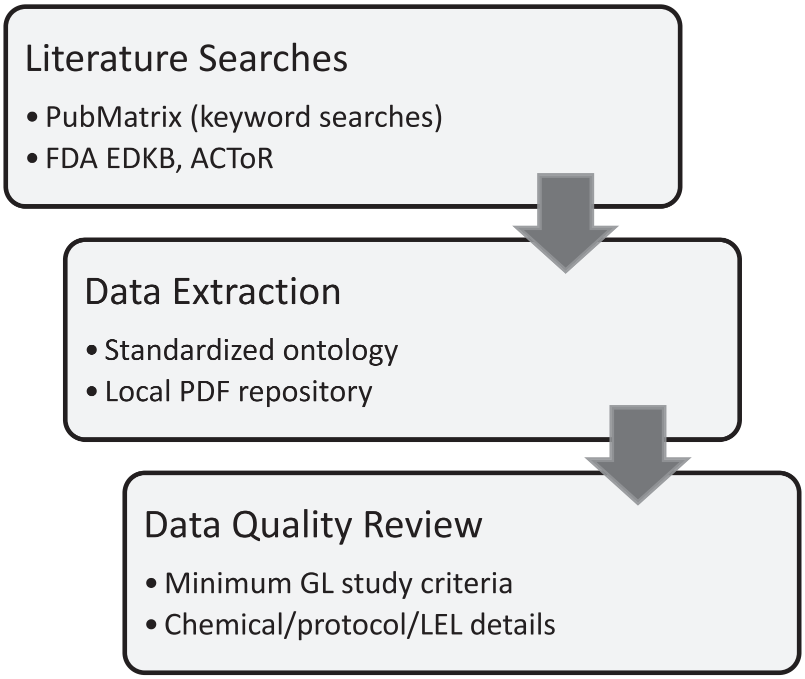

Literature Searches
• PubMatrix (keyword searches)
• FDA EDKB, ACToR
Data Extraction
• Standardized ontology
• Local PDF repository
Data Quality Review
• Minimum GL study criteria
• Chemical/protocol/LEL details

Figure 2

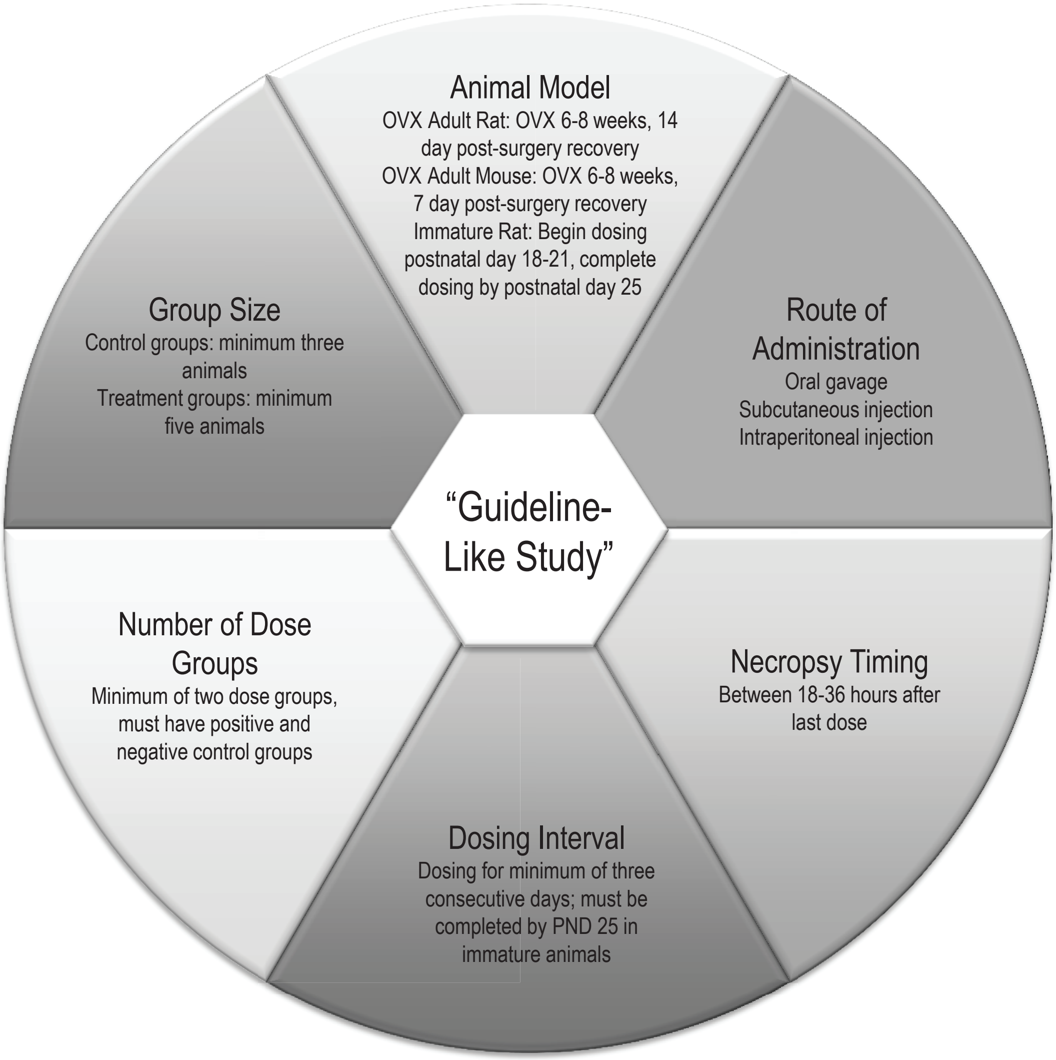

Animal Model
OVX Adult Rat: OVX 6-8 weeks, 14 day post-surgery recovery
OVX Adult Mouse: OVX 6-8 weeks, 7 day post-surgery recovery
Immature Rat: Begin dosing postnatal day 18-21, complete dosing by postnatal day 25
Group Size
Control groups: minimum three animals
Treatment groups: minimum five animals
Route of Administration
Oral gavage
Subcutaneous injection
Intraperitoneal injection
"Guideline-Like Study"
Number of Dose Groups
Minimum of two dose groups, must have positive and negative control groups
Necropsy Timing
Between 18-36 hours after last dose
Dosing Interval
Dosing for minimum of three consecutive days; must be completed by PND 25 in immature animals

Figure 3

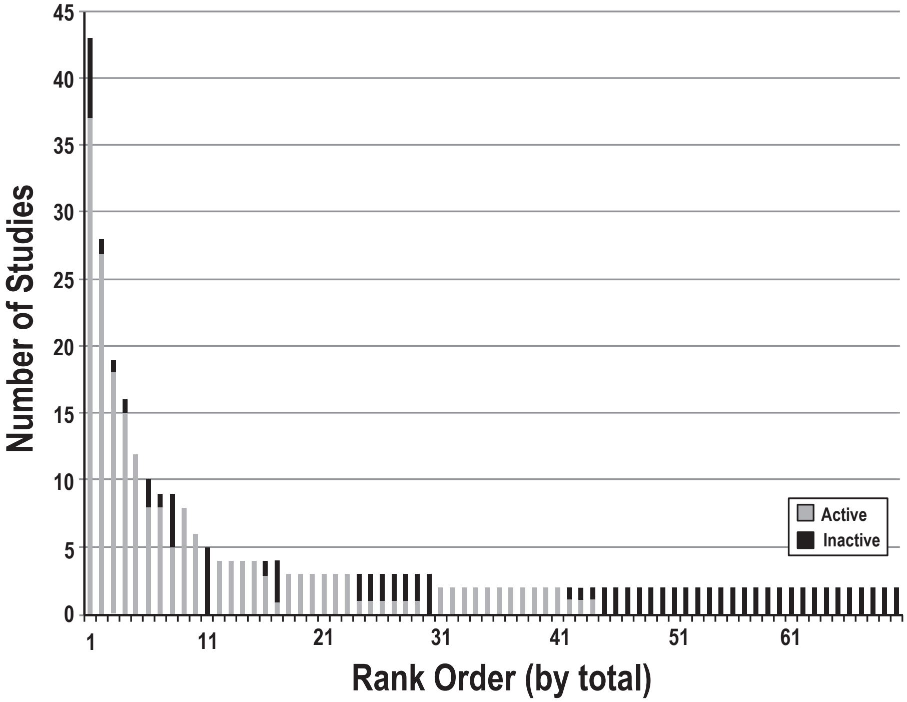

45
40
35
30
25
20
15
10
5
0
Number of Studies
1
11
21
31
41
51
61
Rank Order (by total)
Active
Inactive

Figure 4

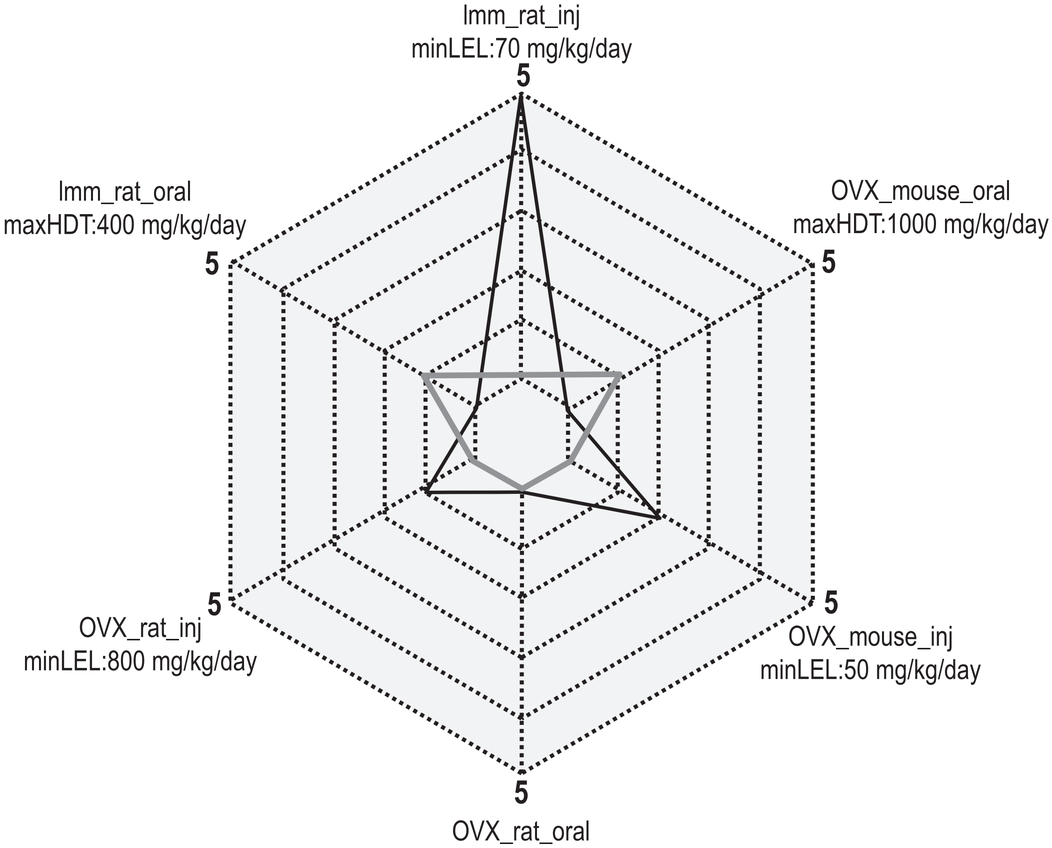
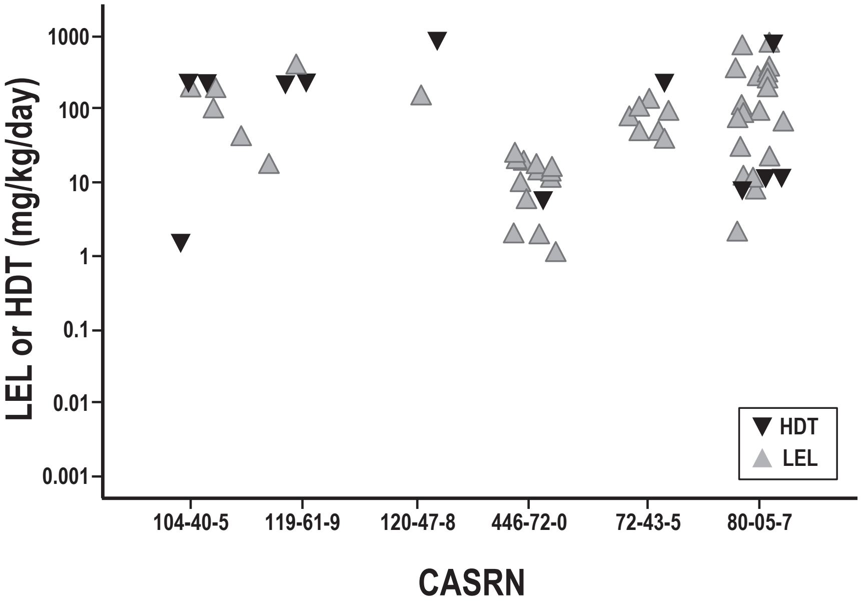

lmm_rat_inj
minLEL:70 mg/kg/day
5
OVX_mouse_oral
maxHDT:1000 mg/kg/day
5
5
OVX_mouse_inj
minLEL:50 mg/kg/day
5
OVX_rat_oral
5
OVX_rat_inj
minLEL:800 mg/kg/day
5
lmm_rat_oral
maxHDT:400 mg/kg/day

Figure 5


LEL or HDT (mg/kg/day)
1000
100
10
1
0.1
0.01
0.001
104-40-5
119-61-9
120-47-8
446-72-0
72-43-5
80-05-7
CASRN
▼ HDT
▲ LEL